ARI Research Note 86-102

(1)

ANALOGICAL DECISION MAKING

Gary A. Klein
Klein Associates

for

Contracting Officer's Representative
Judith Orasanu

BASIC RESEARCH LABORATORY
Milton R. Katz, Director

U. S. Army
Research Institute for the Behavioral and Social Sciences

December 1986

87 4 3 051

U. S. ARMY RESEARCH INSTITUTE

FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency under the Jurisdiction of the

Deputy Chief of Staff for Personnel

EDGAR M. JOHNSON
Technical Director

WM. DARRYL HENDERSON
COL, IN
Commanding

Research accomplished under contract
for the Department of the Army

Klein Associates

Technical review by

Steve Kronheim

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER ARI Research Note 86-102	2. GOVT ACCESSION NO. AD-A178836	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Analogical Decision Making		5. TYPE OF REPORT & PERIOD COVERED Interim Report July 85 - July 86
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Gary A. Klein		8. CONTRACT OR GRANT NUMBER(s) MDA 903-85-C-0327
9. PERFORMING ORGANIZATION NAME AND ADDRESS Klein Associates, P.O. Box 264, Yellow Springs, Ohio 45387		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 2Q161102B74F
11. CONTROLLING OFFICE NAME AND ADDRESS U.S. Army Research Institute for the Behavioral and Social Sciences, 5001 Eisenhower Avenue, Alexandria, VA 22333-5600		12. REPORT DATE December 1986
		13. NUMBER OF PAGES 29
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) - -		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE - -

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

- -

18. SUPPLEMENTARY NOTES

Judith Orasanu, contracting officer's representative.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Decision-Making	Expert-Novice Comparisons
Group Dynamics	Judgement
Firefighting	Uncertainty
Cognition	Decision Point

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The overall goal of the research program this research note describes is to develop a descriptive model of decision making, a model which can be applied to tactical and strategic decision-making domains. Variables such as time pressure, risk, and levels of expertise, will be explored to see how they affect the use of decision strategies and situational assessment. Analytical and recognitional models of decision making will also be contrasted.

(over)

ARI RESEARCH NOTE 86-102

20. Abstract (continued)

The results obtained so far suggest the general applicability of the recognition-primed decision (RPD) model. Even less experienced decision makers don't rely on analytic decision strategies where time pressure is not great, and where group dynamics are operating. Decision makers generally judge a situation to be of a familiar type, generate the most typical reaction, evaluate that reaction, and then put it into effect without further evaluation if it is plausible. Only if it is not plausible, is an attempt made to modify it, and it must be judged implausible for the next most typical action to be generated.

Situation assessment appears to be the most critical stage of naturalistic decision making, and differences in situation assessment appear to account for the primary differences in the quality of decisions made.

TABLE OF CONTENTS

ABSTRACT

The overall goal of this research program is to develop a descriptive model of decision making that can be applied to tactical and strategic decision making domains. We wish to explore variables such as time pressure, risk, and levels of expertise as they affect the use of decision strategies and situational assessment. We also want to contrast analytical and recognitional models of decision making.

Year 1 was devoted to field studies and knowledge elicitation interviews with decision makers including urban fire ground commanders, wildfire incident commanders, and Army Armored Division platoon leaders. In addition a study we performed comparing the quality of chess moves under speeded and non-speeded conditions served as an analogue for real-world strategic and tactical decisions. All of these studies allow a comparison of the decision making at higher- and lower-levels of expertise.

These studies were designed to test the generality of a Recognition-Primed Decision ~~(RPD)~~ model as a function of different domains, varying experience levels to group decision settings and varying levels of time pressure. The results obtained thus far suggest that the model has general applicability. Even less-experienced decision makers, in situations where time pressure is not great and where group dynamics are operating, do <u>not</u> rely on analytic decision strategies. Decision makers most often judge a situation to be familiar, generate the most typical reaction, evaluate that option, and then implement it without further evaluation if it is plausible. If it is not plausible, an attempt is made to modify it. Only after it is judged implausible is the next most typical action generated. Situational assessment appears to be the most critical stage of naturalistic decision making, and differences in situational assessment account for the primary differences in the quality of decisions.

INTRODUCTION

The overall goal of this research program is to develop a descriptive model of decision making in natural settings. Our focus is environments in which strategic and tactical decisions must be made under conditions of extreme uncertainty, risk and time pressure. Because the findings of the prevailing laboratory-based research paradigms have proved difficult to translate into usable programs for such real-world domains, our approach has sought to develop alternative naturalistic and quasi-naturalistic study methods.

The first year studies which are reviewed in this report were designed to test and extend the findings and methods of an earlier investigation of urban fire ground commanders' (FGCs) decisions by urban fire ground commanders (FGC) at the scene of a fire (Klein, Calderwood, & Clinton-Cirocco, 1985). This task parallels some important aspects of a military command and control environment. FGCs must establish fire ground strategy and oversee tactical maneuvering against a potentially lethal and rapidly advancing hazard. Because FGCs in larger urban settings have had extensive command experience, the decision strategies observed in this domain would serve to complement the more prevalent and possibly misleading procedure of observing naive subjects making decisions in artificial and contrived tasks (These criticisms have been made by, e.g., Brandt, 1981; Ebbeson, & Konecni, 1980).

The method used in the Klein et al. (1985) study was a modification of the critical incident technique developed by Flanagan (1954), a semi-structured interview organized around a "critical" or "non-routine" incident (fires and rescues) in which the officer had made command decisions. An officer was asked to recount the incident in his own words and then to construct a detailed timeline of all the important events which he could recall including what he had seen, heard, felt, smelled, or thought at each event time. Each command decision was then probed along a number of dimensions including:

1) the objective, or goal, of the decision;
2) the nature and sources of information relevant to the decision;
3) what other options were considered;

4) how the chosen option was selected (i.e. could a selection rule be articulated, what evaluation dimensions were contrasted);
5) the level of experience or training necessary to make the decision (to interpret the cues or know which cues to look for);
6) the amount of time taken to make the decision;
7) what kinds of critical knowledge or cues were missing.

Thirty-two incidents were collected from 26 officers with an average of 23 years of firefighting experience. For coding purposes, 156 decision points from these incidents were extracted and characterized along the probed dimensions.

Several unexpected results emerged. The earliest and most striking finding was the fact that FGCs would frequently deny that they had made any decisions, at least in the usual sense of a selection among alternative options. That is, in the course of an interview, even when it was clear that a command had altered the "flow" of the incident and there were (by his own admission) other courses of action which could have been taken, an FGC would nonetheless assert that he had not considered other options. In response to our repeated probes about option selection, the officers would frequently say that their actions were "just automatic" or "based on experience" and that there was no time to generate or deliberate between alternative courses of action. While it is possible to maintain that these statements are due to an inadequate memory for decision events, or to some demand characteristics of the method, the consistency and adamancy of these reports convinced us to accept them as having an important degree of validity.

At the very least, it appeared that our definition of decision making developed from an experimental perspective did not match the phenomenological reality of these highly experienced decision makers. For this reason, we had to develop the concept of a "decision point" for the unit of our protocol analysis, a point in time when alternative courses of action were clearly available even if there was no consciousness of deliberation between options. Over 80% of the decision points which were elicited were of this type; a course of action had been implemented without conscious awareness of having deliberated between alternatives. This finding would seem to have important implications for decision support and training programs in which it is assumed that the best decisions require exhaustive option, generation, and analytic evaluation strategies.

Although the majority of the decision points were categorized as non-deliberated, the 20% that had been consciously deliberated also contained some surprising aspects. We had initially hypothesized that, due to time pressure and workload constraints, only two or three options would be considered at any one time and would only be compared on a limited number of evaluation dimensions. We had also expected, on the basis of other work on natural reasoning (Klein, 1980; Klein & Weitzenfeld, 1982), that option selection would frequently be made on the basis of an analogical reasoning process in which the current situation was comparded to a similar, previously encountered situation. Neither of these expectations were borne out by the data.

First, FGCs seldom reported using concurrent option evaluation, in which two or more options would be contrasted on one or more dimensions. Instead, the data suggested a serial decision strategy, in which a single option was generated, and then either implemented or rejected on the basis of a rapid assessment of plausibility. If rejected, the very next (primed) option is examined for plausiblity and implemented or rejected, etc. For example, in an incident where an unconscious woman was suspended on a sign over a freeway, five different methods of rescue were considered (involving different apparatus), yet the officers account indicated that none of the methods were ever directly compared. Rather, each option was evaluated serially, in this case using mental imagery to "test" the outcome of the procedure. The first plausible option was selected and implemented.

We proposed a recogniton primed decision (RPD) model as a description of these data. Simon's (1955) notion of satisficing is consistent with the model, but Simon was concerned with establishing the limits of rationality in decision making. In this domain in which the decision maker must be prepared to act very rapidly, this may be the most rational and effective strategy, and we wish to stress this aspect of expertise. The data did suggest that the less experience a commander had with a type of incident and/or the less time pressure involved, the more likely we were to find evidence of concurrent option comparison strategies.

Secondly, we found very little evidence of the analogical reasoning that we had predicted. In only three cases did an officer report that a particular past event directed him to select a course of action, and in no case was an entire incident treated as an analogue for another. Rather the analogues seemed to serve as "flags", alerting the officer to dynamics of the situation that needed attention. For example, in one case an officer ordered the crowds to be moved out of an area, he reported that he normally would not have taken this action. The action was triggered when he noticed that the building had billboards on its roof and he remembered an incident in which a billboard had collapsed on civilian onlookers.

Perhaps our most important finding, and one we also had not foreseen, emerged from our difficulty in coding decision points independently of the events surrounding them. That is, each decision point could only be understood in terms of the goals, perceptual cues and knowledge that were operating at the time of the decision. We therefore had to find ways of tracking the officers' "situational assessment" throughout the course of the incident, and it now appears to us that situational assessment is crucial to understanding time-pressured decision making.

In fact it should be noted that the critical decision method was somewhat misnamed. Many of the decisions probed were routine, it was the situational assessments which most often defined a case as 'non-routine' and challenging. In some cases, the initial situational assessment was maintained throughout the incident, but more often there were from three to five elaborations or shifts, defined in terms of cues which affected the understanding of the causal dynamics of the case and resulted in changes in a primary or secondary goals. One of our present goals is to find means of representing a situational assessment that can aid our understanding of this important component of decision making.

In summary, the difficulties we had in deriving a satisfactory definition of "decision" in this study was a revealing and important finding. The literature on analytic decision making based in laboratory explorations had not prepared us for the complex nature of the data we were obtaining. Because we learned so much beyond our original expectations we were convinced that this type of investigation could enhance our understanding of decision making. However, we knew that our exploratory method would have to be refined and validated before the rich source of hypotheses suggested by our protocols could be tested or achieve any practical success. Each of the four studies proposed and carried out in Year 1 of this effort addressed methodological and/or theoretical issues raised by these findings.

OVERVIEW OF YEAR 1 RESEARCH

Five studies were initiated in Year 1 of this three-year study:

Study 1 -- Critical Decision Methodology with Israeli Command, Control and Communication Personnel (Israeli Officers).

Study 2 -- Critical Decision Methodology with Wild Fire Commanders (Wildfire).

Study 3 -- Recognitional and Calculational Capacities as a Function of Skill Level in Chess (Chess).

Study 4 -- Critical Decision Methodology with Army Armored Division Commanders (Tank Commanders).

Study 5 -- Critical Decision Methodology to Contrast Skill Levels of Urban Firefighters (Expert-Novice).

This was more than we could expect to finish in a single year, but rather than sacrifice any of the opportunities which we pursued, we chose to gather as much momentum as possible and adjust our second year planning on the basis of the range of our preliminary findings. Except for Study 1 (Israeli Officers), all of these studies have progressed at or beyond our initial expectations in terms of the wealth of data which we have obtained. Studies 3 (Wildfire), 4 (Tank Commanders) and 5 (Expert-Novice) were carried out so that they have already provided data and organizational structure for Year 2 investigations.

A major goal for Year 1 was to refine the critical decision method and to explore means for adapting the method to different content domains. Because Study 5 (Expert-Novice) was most similar to the original firefighter study, our major efforts toward refining the interviewing and coding methods were started here. The interview guide was subjected to four major revisions. The progression of these revisions was generally to become more and more structured in our efforts to increase the objectivity and reliability of our coding. Interestingly, we found that increasing the degree of structure during an interview did not necessarily guarantee better data, in fact we reached a point where our data began to suffer. As our conceptualization of our data became clearer, we found that we could more comfortably give "control" of the interview back to the decision maker and still elicit the desired information. Part of the learning process was also involved in how to teach the method to new interviewers who would be involved in these and future studies. At this point we feel that our efforts have paid off nicely. From our initial loosely structured attempts to get firefighters to "tell us about it" we have begun to accumulate a pool of techniques and types of probes that elicit detailed, specific information about the conscious processes of real-world decision making. Each study has improved the method by testing specific techniques and retaining those found to be useful. Another goal was to test and extend our findings and to compare more- and less-experienced decision makers. We feel that progress towards these goals have been excellent.

Study 4 (Tank Commanders) has been an especially exciting project as decision points could be elicited from both highly experienced instructors and their students <u>for the same exercise</u>. This has allowed a direct comparison of decision strategies not possible in our other studies and holds the promise of being the most detailed and systematic comparison of situational assessment for "experts" and "novices" that we know of. It has also allowed us to incorporate on-site observation and immediate recall into the retrospective critical decision method.

Study 2 (Wildfire) also allowed on-site observation and immediate probing and is our first chance to examine group and distributed decision processes. Preliminary results suggest that there was not as great a use of analytic option generation and concurrent evaluation as we had expected. Thus, the RPD strategy appears to be more general than we anticipated. The same is true for the Expert-Novice comparison of Study 5. Although the types of decisions made by more- and less-experienced commanders are quite different, because the span of control of less-experienced officers tends to be less, it does not appear that the decision strategies themselves are different.

Study 3 (Chess) offers independent support for the importance of rapid recognitional processes in expert decision performance, and we hope to carry out additional research in Year 2 which will directly manipulate situational assessment in this domain.

A more detailed description of the goals and progress of each of the studies follows.

Study 1: Critical Decision Methodology with Israeli Command, Control, and Communication Personnel (Israeli Officers)

The study in Israel was planned as an investigation of command and control decision making, relying on retrospective protocols for actual operations. The subjects were to be Israeli Defense Force personnel. At present, we have encountered difficulties in the design of this project. The arrangements requested by the IDF may compromise some of our data-gathering procedures to such an extend that the data we gather would not be sufficiently rich to justify the study. We are currently engaged in discussions in order to clarify the design. It is possible that this study will be deleted and replaced with another research investigation that is acceptable to the Contract Monitor.

Study 2: Critical Decision Methodology with Wildfire Commanders (Wildfire)

The purpose of this study is twofold. First, we want to compare the decision making strategies of experts in a situation where time is not always a critical element to the strategies employed by the urban firefighter whose decision making strategies are governed by the pressure of time. Second, we wish to explore the strategies employed in team decision making and distributed decision making in order to enhance our understanding of the entire decision making processes in complex organizations.

A trip was made in March to the Boise Interagency Fire Control Center to pave the way for gathering data on a wildland fire during the summer. During our visit we learned about the Incident Command System which is the organizational and command structure imposed upon the fire suppression activities of major wildland fires. We obtained sufficient familiarity about fire suppression activities to enable us to understand what we would see during the summer when observing actual suppression activities. We set up the process whereby we would be attached to an Overhead Team (one of 17 standing teams across the country) and be called to a fire when that team was activated during the fire season.

On August 12, we received the call to the Garden Valley Complex Fire near Boise, Idaho. On August 13, two trained interviewers joined the Overhead Team in Boise. These interviewers spent eight days on the fire, living with the firefighters, observing their operations, and conducting interviews when time permitted.

We had planned to contrast the decision making strategies among the various levels of expertise and functions within the team. However, additional research opportunities were afforded. This complex of fires necessitated a more intricate organizational strategy than is usually the case in wildland fire suppression. Five separate fires burned during our stay there, each with its own team of firefighters. Several of the smaller fires were staffed by Class II teams, a large fire by a Class I team, and the overall coordination of firefighting activities performed by an additional Class I team. The difference between Class I and Class II teams is one of experience and expertise. The members of the Class I team have more years of experience fighting fires and have received more formal and informal training in fire suppression and accompanying management activities. We, therefore, could sample several levels of expertise <u>between</u> teams as well within teams. Additionally, the presence of the Class I team that coordinated the firefighting activities (called, Area Command), gave us the opportunity to see decision making at a level one step removed from the supression activity itself. We hope to be able to add this layer of complexity to our analysis of team and distributed decision making.

We made extensive use of tape recorders and notes documenting the decision making processes we observed on the fire. At this time, we have data obtained from approximately 15 firefigters of varying levels of expertise. We are anticipating that we will have around 70 decision points. In addition, we will be collecting retrospective data in mid- October. These interviews will be aimed at eliciting from these experts their assessment of the most critical decisions in which they participated during the fire, thus adding another element of interest to our study. Additionally, we will have the opportunity to reexamine some of our decision points and to test the reliability of recall of these experts.

<u>Study 3</u>: <u>Recognitional and Calculational Capacities as a Function of Skill Level in Chess</u> (<u>Chess</u>)

The recognition-based decision making we have found so prevalent in our other studies would seem to require fairly extensive experience for the necessary perceptual learning and development of situational assessment to have occurred. However, although the distinction between recognitional and calculational components of decision making appears to be an important one (Hammond et al., 1984) methods for examining these two aspects of performance need to be developed.

In this study, we proposed a method based on the assumption that one difference between recognition and calculation is the relative duration of the two processes. Calculational components of decisions (i.e. identifying options, comparing and evaluating alternatives) constitute information processing sequences that are inherently time-dependent. When time constraints become severe, stages within the overall sequence are omitted or truncated, and performance suffers. In contrast, the holistic perceptual nature of recognitional processes leaves them relatively time-independent. For example, recognition of a familiar face is virtually immediate, regardless of the number of faces known.

A chess study begun some time ago at Klein Associates (Klein & Peio, 1982) seemed relevant to these various issues and led us to a renewed interest in pursuing this work. Briefly, the Klein & Peio study examined the effect of time pressure on the decision strategies employed in actual chess games, using two levels of skilled players (Masters and Experts, as ranked by the U. S. Chess Federation), and under conditions of regulation play (a minimum of 50 moves in two hours) and speeded play (5 minutes total time for each player). If it is the case that highly-skilled players rely more heavily on their recognitional capabilities in selecting a move than do less-skilled players, then time limitations should have less impact on the quality of their play. Conversely, if less expert players are more likely to calculate move consequences, then they should evidence greater impairment in the quality of their games under the time-pressured conditions of speeded play. Thus, to the extent that differences in expertise reflect differential use of recognitional processes, then discrepancies in the quality of players' games should be greater under speeded play conditions than under regulation play conditions. But if players differ primarily in their calculational abilities, then the greater amount of time available to select moves provided by regulation play should produce a greater divergence between the quality of the two groups' games than would occur under speeded conditions.

Although the method used in this earlier study appeared promising we failed to reach statistical significance for the predicted interaction. It seemed possible that our failure to support the recognitional hypothesis may have been due to the fact that the skill levels employed in that study differed by only a single rank, and are both highly skilled in absolute terms. Thus, we decided to gather additional data from Class A chess players (one rank below Experts) to see if this degree of differentiation in skill levels might allow the hypothesized effects to emerge. Accordingly, three Class A players were located and a double round robin tournament played, yielding 6 speeded and six regulation games.

We were extremely fortunate to have obtained the services of two International Grandmasters to rate these games and to re-rate the games from the previous study as well. Our Grandmaster raters evaluated each move contained in a game for degree of complexity and rated its quality on a five-point scale. These assessments allowed us to separate obvious, straightforward simple moves from more challanging (complex) ones, and to examine the quality of moves within each of these categories.

Analysis of variance of these measures revealed no significant effects of player skill (Master, Expert, Class A) and type of game (Speeded, Regulation) on quality of play for simple moves. However, when the proportion of complex moves that had been rated "poor" (13 or less) was examined, the hypothesized interaction of skill level and game speed did emerge ($F(11,9) = 8.60$; $p < .02$). In speeded play, Class A players had produced complex moves of significently poorer quality as compared to games played by the other skill groups.
We have re-examined our original rating scale and believe that it was insensitive to some of the key differences were trying to find. One of the Grandmaster coders who had participated in the study has made a number of suggestions for altering the rating scale that he felt would provide both increased variability and a more reliable index of move quality. He indicated his willingness to re-rate the games, using the

new scale, and we therefore decided to have the Master and Class A games rescored. In the interest of time and money, it was decided not to have the Expert games rescored, since they had differed little from Master level games. The games are currently being re-rated and we expect the analysis to be complete by mid-October.

Study 4: Critical Decision Methodology with Armored Division Platoon Leaders and Company Commanders (Tank)

This study was designed to examine the decisions of tank platoon leaders and company commanders during training exercises at Ft. Knox, Kentucky. The exercises were tracked from early training stages through part of the force-on-force maneuvers at the end of a "seven-day war" simulation. Thus, this study afforded the opportunity to verify and anchor decision points by performing the critical decision interviews in conjunction with on-site observations. It also allowed a unique opportunity to compare the situational assessments and decisions strategies of more- and less-experienced personnel performing the same decision task.

After an initial planning and information gathering trip, a pilot study was performed in early May that permitted us to observe the training process and to modify the knowledge elicitation method. Several limitations had to be accomodated, primarily the limited amount of time between exercises in which we were able to interview personnel. An approach was developed that allowed us to touch on all relevant material in five minutes, yet could be expanded to use whatever additional time was available. Our direct observations were a distinct advantage in the knowledge elicitation method, because the interviewee did not have to reconstruct the incident for the interviewer before critical decisions could be probed.

The first data-collection trip was carried out in mid-May and involved direct, simultaneous observation and interviews with two individuals per exercise: a highly-experienced trainer (TCI) and a trainee (AOB). Both individuals experienced the same training exercise from the same tank. One researcher rode in the tank with the TCI and AOB. The second researcher observed the training exercise from the vantage point of the company commander's jeep, where the scenarios for the exercise were being developed. This second researcher was able to observe the "logic" of the exercise (i.e., determine what the company commander was hoping the AOB would do and learn during the exercise).

Immediately following each exercise, a rank-ordered list of the "most difficult/challanging decisions" was obtained from the TCI. Both the TCI and the AOB were independently interviewed immediately after the exercise. Approximately 35 separate decision points were gathered from 6 TCIs and 16 AOBs.

Information from this initial data gathering phase is presently being coded and analyzed. Our measures include various aspects of situational assessment, options available/generated, goals, basis for option selection, etc., from the dual perspective of TCI and AOB. With 90% of these decision points analyzed, it appears that the most enlightening aspect of this study may be the demonstration of differences in the situational assessment of the TCIs and AOBs. Despite the fact that experts and novices in this study were engaged in the same task, at close physical proximity to one another, their reports of what they considered to be salient aspects of the situational surround differ, often dramatically. We have found the present study extremely productive for developing and refining measures of situational assessment. We have planned the second data-collection trip, scheduled for early October 1986, to further explore these differences.

Study 5: Critical Decision Methodology to Contrast Skill Levels (Expert/Novice)

Chronologically, this was the first study initiated in Year 1 and it was designed to address several of the conceptual and methodological issues raised by our initial examination of decision making on the fireground (Klein et al., 1985). Specifically, we sought to replicate the findings of the earlier study and to refine our knowledge elicitation and coding procedures. In addition, we wanted to examine how individuals with varying degrees of experience in a domain differ in their decision making. On the basis of a model of expertise proposed by Dreyfus and Dreyfus (1986) we predicted that we would find more evidence of rule-based and analytic decision strategies for less-experienced officers than for more highly-experienced officers. We also expected to find a higher incidence of analogical reasoning (using a specific remembered event as the basis for a decision) among the less-experienced officers. We had speculated that our failure to find more evidence of analogical decisions in the earlier study was due to the fact that experienced officers had had so many similar experiences that these had become merged into a "prototype" and would not be accessible as a single remembered event.

We gained the cooperation of six professional urban fire departments and asked department chiefs to identify officers from among their most- and least-experienced commanders. We interviewed twelve "experts" who had the rank of captain or above and twelve "novice" officers who were newly promoted lieutenants and had at least one fire ground command to their credit. Because of the different promotion opportunities and incident base rates in the different departments, rank alone or years of experience were inadequate measures of level of expertise. For example, a seven year veteran in a very busy company may have the equivalent experience of a fourteen year veteran in a company where base rates are low or fires tend to be small and routine. By interviewing the same number of experts and novices in each department, we hoped to control for some of these extraneous factors in making the expert-novice comparisons. In addition, details of each officer's experience and base rates obtained from each department will allow us to estimate the absolute experience level of each officer to some degree.

One of the outcomes of this study is in relation to the explicit goal of refining the critical decision method in ways that would increase the objectivity of the decision point coding. In the earlier study, decision points often had to be characterized inferentially from an overall understanding of the incident account. Often a decision point would have been "missed" during the interview, so that it was not explicitly probed. This resulted in ambiguities in the coding and/or lost data if inter-coder disagreements could not be resolved by referring to the taped interviews. Our goal was to decrease the amount of time required to code an interview and increase the inter-coder agreement by defining and probing each decision point in the interview in a more standard way.

Modifications of the interview guide designed to increase coding reliability were tested in preliminary interviews. Several techniques were found to be useful, although not all attempts to increase the interview structure were judged to be worthwhile. For example, when each decision point was probed in a totally standard way, the interviews became repetitive and boring to the interviewee and substantially increased the length of the interviews without noticably enhancing the quality of the reports. We became convinced that a semi-structured format in which we learned to be sensitive and active listeners was the best approach.

The primary modification of the method that was incorporated after preliminary interviews was the use of a "consensus timeline". This was simply a timeline which was summarized from the interviewers individual timelines and shown to and verified with the officer. This allowed him to modify our understanding and structure of the incident and fill in details or make corrections before decision probing began. The officer verified the number and description of each decision point by indicating a "yes" to one of several probes which substantiated that he had either consciously deliberated between alternatives at a point in time or that reasonable alternatives were clearly available. In either case, other options were elicited from the officer and, if considered, the reasons for acceptance and rejection were probed.

Coding of the Klein et al. (1985) study had been largely narrative, so another methodological goal of the present study was to create a coding scheme which would facilitate analysis and allow aspects of the data to be computerized. A classification system was developed that contains some 48 categorical variables. Thirty of these pertain to the overall incident and interview charactertistics, for example the type of incident, the experience level of the officer, the command level of the officer during the incident, why the incident was chosen, etc. Many of these factors will have to be considered in trying to interpret any differences in observed decision strategies.

Over 80% of the coding has been completed and summaries should be completed by the end of September. It appears that our original hypotheses will have to be rejected. We have not found that novices are using context-free rules to a greater extend than experts, or that they are relying on analogues as the basis for option selection. Rather, novices and experts appear to use identical strategies in making decisions. Differences are more related to the breadth of factors considered, with experts having to consider a "bigger picture." The most problematic aspect of the design, in fact, is that the experts were more frequently in charge of larger, more complex fires, and that the novices were more likely to be making tactical rather than strategic decisions. The most striking difference observed thus far is that experts are more likely to make a decision based on a judgment of a future contingency, taking into account "worst case" changes in the scenario, whereas novices are reacting more to immediate situation states.

REVIEW OF FINDINGS

At this point in the project we have collected a great deal of data but we have not yet completed any of the analyses. Nevertheless, we have reached some tentative conclusions about factors affecting decision making. In this section we will review our primary conclusions based on our findings so far.

The study on chess performance was our only study not based on the critical decision method. Under different degrees of time pressure, we compared differences between more- and less-skilled chess players. The recognitional model predicts that chess Masters and Class A players would differ most radically under speeded conditions, whereas a calculational model would assert that the more time available for calculation then the greater the difference between Masters and Class A players. Thus each model predicts an alternative interaction. Preliminary results from this study support the recognitional model in that Class A players showed greater performance decrements, measured by moved rates 'poor', than Masters. We see these findings as important because they indicate that highly skilled people can maintain high quality performance in complex decision making under extreme time pressure. These results demonstrate the strengths of human expertise, rather than the biases that have become the focus of decision research. Further analyses are planned after games have been re-rated using a potentially more valid and reliable scale.

For the three critical decision studies that we were successful in carrying out, our ongoing analyses are suggesting several important conclusions. We have summarized what we have learned in Table 1 in terms of our current questioning of beliefs that we, and others, have held about decision making. Each of these beliefs, and our data relevant to it, will next be discussed in turn.

TABLE 1

QUESTIONABLE BELIEFS ABOUT DECISION MAKING

1. Subjects should generate as many alternatives as possible in making decisions.
2. Good decision making requires analytical deliberation between options, at least if there is sufficient time.
3. Group decision making will increase the likelihood of cases in which there is an analytical deliberation between options.
4. In the process of selecting an option, proficient decision makers match new situations to analogues and prototypes stored in memory.
5. Identified decision biases, such as availability and representativeness, result in poor decisions.
6. Experts are sensitive to context whereas novices follow context-free rules.
7. Experts are sensitive to deep features of a problem whereas novices deal only with surface features.
8. Context is a central concept for understanding proficient decision making.
9. The key to decision making is the reduction of uncertainty.

---

1. Subjects should generate as many alternatives as possible in making decisions. Our studies of urban firefighters, and of tank commanders as well, have revealed virtually no cases where the decision maker sought to generate an exhaustive set of options. In fact, there were few cases where more than two or three options were generated, and these were in cases where time pressure and expertise tended to be low.

We believe that this is because any attempt to generate a larger option set would have carried with it the cognitive burden of evaluating these options. The time pressure of these situations is not conducive to such a strategy. Indeed, generation of options would be counter-productive in the situations we studied, and perhaps in all dynamic situations because there is a good chance that conditions will change and opportunities will be missed by the time the set of options is generated and evaluated.

2. Good decision making requires analytical deliberation between options, at least if there is sufficient time. We had originally speculated that our failure to find analytic option generation and evaluation in the study of urban firefighting (Klein et al., 1985) was due to the extreme time pressure involved (most decisions were made in less than one minute). However, in the Wild- fire study, where incidents continue for days and each decision is less time constrained, we have still found that in the majority of cases only one option was generated and evaluated at a time. In a study of design engineering decisions using the critical decision method (Klein & Brezovic, 1986), time pressure was even less severe, yet still the serial evaluation strategy was very common.

It is beginning to appear to us that the best way to get decision makers to generate large numbers of options and compare them to one another is to make this a formal requirement of the task. Left to their own devices, decision makers will tend to rely on serial generation and evaluation of one option at a time. The fact that concurrent evaluation appears to be used so rarely in everyday decision making should make us pause before attempting to implement decision support procedures which forces this mode on an operator.

3. Group decision making will increase the likelihood of cases in which there is analytical deliberation between options. We had thought that our RPD model was applicable to individual decision making, but that in group settings there would typically be several options generated by different group members. We anticipated that there would also be pressure to systematically evaluate these options in order to arrive at one option as superior to all others. Thus far, our study of group decisions (Wildfire) has revealed the same pattern of evaluating one option at a time found for individual decision makers. Although there are a greater number of cases in which several options were generated and analyzed, these are still in the minority.

What seems to happen is that once a situation is identified, there is a quick recognition of a likely way to proceed and the attention of the group turns to problem solving. Group members try to identify possible pitfalls and to find ways around these. If too many potential pitfalls are found, then the option is abandoned and another one generated in its stead.

It was our impression in observing these decision processes, that the serial evaluation strategy was a productive activity that furthered group cohesiveness. In contrast, the standard laboratory decision making paradigm where one option is to be selected from several, may be a divisive activity that would diminish group morale.

4. <u>In the process of selecting an option, proficient decision makers match new situations to analogues and prototypes in memory</u>. Despite our expectations and hardest probing, we have found little evidence of the use of specific analogues as a basis for selecting among options. Furthermore, our subjects show no evidence of having prototypes that they have developed and stored for the future. We still believe that skillful decision making depends on recognitional matches and judgments of prototypicality. It may be that episodic memories are accessed and synthesized on the spot in order to generate the judgments of typicality. A similar hypthesis has been offered by Kahneman and Miller (1986), and is more or less implicit in the work of Lakoff (1986) and of Jacoby and Brooks (1984).

5. <u>Identified decision biases, such as availability and representativeness, result in poor decisions</u>. Until very recently, a focus of decision research has been demonstrations of how human decision making is biased relative to normative standards (e.g., Slovic & Lichtenstein, 1971; Tversky & Kahneman, 1974). We start with the opposite assumption; that human decision making has evolved to meet the requirements and constraints of real-world tasks. "Biases" such as availability and representativeness reveal the fact that skilled decision makers have learned to rely on their episodic memory. They rely on prior experience. The only way to free a system from these biases would be to eliminate episodic memory, and to eliminate judgments of typicality. We believe this would also eliminate much of human expertise.

If one is intent on having a 'perfect' decision system that makes no errors, then decision biases are a real concern. However, it is important to recognize that the cost of eliminating these biases is to reduce the range of decision situations that the system can address. Human decision makers appear to have the ability to flexibly handle a great variety of situations at only a small cost in accuracy. In addition, humans can handle ill-defined domains that formal analytical methods cannot address.

We believe that casting analytical decision making (e.g. relying on Multi-Attribute Utility theory, Bayes' theorem, etc.) as a prescriptive goal is misguided. These strong analytical techniques work best in artificial laboratory conditions. In many naturalistic settings it is not clear how to even begin to apply them, let alone expect that human decision makers would be able to use them to advantage. Our research has yet to demonstrate cases where these analytical methods would be useful. We sense that others in the field of decision research are also beginning to swing away from these analytical methods as prescriptive models.

6. Experts are sensitive to context whereas novices follow context-free rules. It has been suggested (e.g. Dreyfus & Dreyfus, 1986) that experts are sensitive to the context but that novices are not and must therefore rely on context-free rules. We no longer view context as a critial explanatory construct for understanding proficient versus non-proficient decision making. Clearly, experts are sensitive to a greater variety of contextual features. For example, they understand the causal implications of a more extensive and varied set of factors than do novices. Nevertheless, every task occurs within a context and we find both novices and experts applying whatever degree of contextual sensitivity they possess to the task at hand. In our view, to assert that experts are better problem solvers and decision makers because they are more sensitive to context does not seem to provide additional leverage to understanding proficient performance. What does seem important is to explore the specific contextual elements that experts notice that novices do not, and the types of relationships each can infer.

In sum, we are suggesting that experts and novices are not qualitatively different in their general approaches to a task. What does distinguish experts is that they are people who have been through the same or a similar situation so frequently that most, if not all, features of the situation have become familiar. When experts encounter a new situation, they do not function as an expert. For this reason, perhaps we should characterize expertise as a mode of reaction at a given task level (holistic recognition versus analysis of independent features). Experts are people who are able to maintain holistic recognition more frequently and for a wider array of task conditions.

7. Experts are sensitive to deep features of a problem whereas novices deal only with surface features. Gentner (1986) has made this suggestion, and we were originally disposed to accept it. However, in reviewing our data we found that we could not distinguish between deep and surface features. For a given fire, the type of material used for external construction is a surface feature. However, a novice could easily miss its significance whereas an expert would notice its effect on the speed with which the house would burn. So a "surface" feature is one that has no significant causal implications, not one that is concerned only with physical appearance. Our protocols show no evidence that novices have a greater tendency to focus on surface features. In fact, novices may assert incorrect deep relationships, but they are searching for causal relationships nonetheless.

8. The key to decision making is the reduction of uncertainty. Wohl (1979) has relied on an information-theoretic framework to generate this assertion, and we have found it appealing but ultimately limited.

First, there is a cost to reducing uncertainty. It takes time. Often a decision maker will have to act before all uncertainty is reduced to zero. Therefore there are goals more important than uncertainty reduction.

Second, there are times when decision makers seek to increase uncertainty. If you believe you can handle complexity more effectively than your opponent, you might seek to increase uncertainty.

Third, a more pertinent issue may be situation assessment. Until you are comfortable with your degree of situation asssessment you will seek to acquire additional information, but once you are satisfied you will turn your attention to other aspects of the task at hand.

Fourth, decision support systems that ignore the decision maker's situation assessment and continually attempt to reduce uncertainty can be disruptive. If you have learned all that you need and you are still deluged with additional data, your performance can be degraded.

YEAR 2 PLANS

The immediate plans for Year 2 are to complete all work begun in Year 1 and to modify or replace the Israeli Officer (Study 1) that was not successfully initiated. For Study 2 (Wildfire), this includes the collection in mid-October of retrospective critical decisions, further coding, and data summary. Study 3 (Chess) is currently being re-analyzed and the final report written pending completed analysis. Study 4 (Tank Commanders) coding is almost completed and a second data-collection trip is scheduled for early October. Study 5 (Expert-Novice) is nearly complete and the final report draft is scheduled for mid-October.

After all of the Year 1 studies have been completed, we feel we will have a unique opportunity to synthesize findings from an extensive database of naturalistic decision making. We would like to develop a decision taxonomy based on these data and merge our findings into an overall descriptive framework. Conceptual work on the taxonomy was begun in another project (Klein, Calderwood, & Eggleston, 1986) and is being refined as part of the on-going analyses.

Although the Year 1 studies were carried out with a focus on decision strategies, we believe the data has the potential to greatly increase our understanding of situational assessment as an important component of the decision process. Very little work has been done on "pre-decision" processes (Gettys, 1983) and we believe these processes will prove to be most critical for decision support and training. Several methods of representing situational assessment in our studies is being explored as part of our coding, and these methods will be critically compared before proposing the details of a Year 2 situational assessment study. One representational method allows shifts and elaborations in situational assessment to be tracked throughout an incident based on shifts between hierarchical goals. Another method uses a dendogram map for the critical cues leading to a decision as a means of representing the structure of an individual's situational assessment.

Another set of planned studies will focus on the judgments of typicality that appear to be so important in real-world decision making. We are exploring methods of cluster analysis for uncovering the structure of situational assessment in these incidents. Both the tank force-on-force maneuvers and fire ground training scenarios which have been explored could provide ways of demonstrating the effects of judged typicality on decision processes. Details for conducting a fire ground incident simulation which would allow a direct comparison of the prototypes of experts and novices are being proposed.

Because the serial decision strategy uncovered by our findings has, to our knowledge, not previously been documented, we feel this aspect warrants further study. Because we have not uncovered extensive use of alternative concurrent evaluation strategies in the domains we are studying, we will propose that factors promoting the use of these strategies be studied in more controlled settings.

REFERENCES

Brandt, R. M. (1981). Studying behavior in natural settings. NY: University Press of America.

Dreyfus, H., & Dreyfus, S. (1986). Mind over machine: The power of human intuitive expertise in the era of the computer. NY: Free Press.

Ebbesen, E. B., & Konecni, V. J. (1980). On the external validity of decision-making research: What do we know about decisions in the real world. In T. S. Wallsten (Ed.), Cognitive processes in choice and decision behavior (pp. 21-45). Hillsdale, NJ: Lawrence Erlbaum Associates.

Gentner, D. (1986). Towards a micro-structured account of human reasoning. Paper presented at the Allerton Workshop on Similarity and Analogical Reasoning. Monticello, Illinois.

Gettys, C. F. (1983). Research and theory on predecision processes (TR 11-30-83). Norman, OK: University of Oklahoma, Decision Processes Laboratory.

Hammond, K. R., Hamm, R. M., Grassia, J., & Pearson, T. (1984). The relative efficacy of intuitive and analytical cognition. Boulder, Colorado: Center for Research on Judgment and Policy.

Jacoby, L. L., & Brooks, L. R. (1984). Nonanalytic cognition: Memory, perception, and concept learning. In G. H. Bower (Ed.), The psychology of learning and motivation: Advances in research and theory, 18, 1-47.

Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. Psychological Review, 93, 2, 136-153.

Klein, G. A. (1980). Automated aids for the proficient decision maker. Proceedings of the International Conference of Cybernetics and Society (pp. 301-304).

Klein, G. A., & Brezovic, C. P. (1986). Design engineers and the design process: Decision strategies and human factors literature. (Subcontract for MacAulay-Brown, Inc., U. S. Army Research Institute, Contract No. F33657-82-C-0513).

Klein, G. A., Calderwood, R., & Clinton-Cirocco, A. (1985). Rapid decision making on the fire ground. (Technical Report No. TR-85-46-12), Washington, DC: U. S. Army Research Institute.

Klein, G. A., Calderwood, R., & Eggleston, R. G. (1986). _A preliminary assessment of factors affecting decision complexity_. (Armstrong Aerospace Medical Research Laboratory, Contract No. F33615-82-C-0513). Wright Patterson Air Force Base, Ohio.

Klein, G. A., & Peio, K. J. (1982). The use of analogues in comparability analysis. Paper prepared for Honeywell, Inc.

Klein, G. A., & Weitzenfeld, J. (1982). The use of analogues in comparability analysis. _Applied Ergonomics_, _13_, 99-104.

Lakoff, G. (1986). _Women, fire, and dangerous things: What categories reveal about the mind_. Illinois: University of Chicago Press.

Simon, H. A. (1955). A behavioral model of rational choice. _Quarterly Journal of Economics_, _69_, 99-118.

Slovic, P., & Lichtenstein, S. (1971). Comparison of Bayesian and regression approaches to the study of information processing in judgment. _Organizational Behavior and Human Performance_, _6_, 649-744.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. _Science_, _185_, 1124-1131.

Wohl, J. G. (1979). _Battle management decisions in Air Force tactical command and control_. MA: MITRE Corporation.